

Page 1 of 8

1600

TECH CENTER 1600/2900

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/765,272A

DATE: 02/20/2003 TIME: 16:02:52

Input Set : A:\340P2C2 Substitute Sequence List 14 Feb 2003.txt

Output Set: N:\CRF4\02202003\I765272A.raw

SEQUENCE LISTING

5 (1) GENERAL INFORMATION:

	J	(I) ODNO	THI ORDITION.
	7	(i)	APPLICANT: Choi et. al.
	9	(ii)	TITLE OF INVENTION: Streptococcus pneumoniae Antigens and
	10		Vaccines
	12	(iii)	NUMBER OF SEQUENCES: 454
	14	(iv)	CORRESPONDENCE ADDRESS:
	16		(A) ADDRESSEE: Human Genome Sciences, Inc.
	17		(B) STREET: 9410 Key West Avenue
	18		(C) CITY: Rockville
	19		(D) STATE: Maryland
	20		(E) COUNTRY: USA
	21		(E) COUNTRY: USA (F) ZIP: 20850 COMPUTER READABLE FORM: ENTERED
	24	(v)	
	26		(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
	27		(B) COMPUTER: Dell Latitude C610
	28		(C) OPERATING SYSTEM: Windows 2000
	29		(D) SOFTWARE: ASCII Text
	32	(vi)	CURRENT APPLICATION DATA:
C>	34		(A) APPLICATION NUMBER: US/09/765,272A
C>	35		(B) FILING DATE: 22-Jan-2001
	36		(C) CLASSIFICATION:
	39	(vii)	PRIOR APPLICATION DATA:
	41		(A) APPLICATION NUMBER: 08/961,083
	42		(B) FILING DATE: OCT-30-1997
	45	(viii)	ATTORNEY/AGENT INFORMATION:
	47		(A) NAME: Lin J. Hymel
	48		(B) REGISTRATION NUMBER: 45,414
	49		(C) REFERENCE/DOCKET NUMBER: PB340P2C2
	52	(ix)	TELECOMMUNICATION INFORMATION:
	54		(A) TELEPHONE: (301) 610-5790
	55		(B) TELEFAX: (301) 309-8439
	58		RMATION FOR SEQ ID NO: 1:
	60	(i)	SEQUENCE CHARACTERISTICS:
	61		(A) LENGTH: 1999 base pairs
•	62		(B) TYPE: nucleic acid
*	63		(C) STRANDEDNESS: double
	64		(D) TOPOLOGY: linear
	67		SEQUENCE DESCRIPTION: SEQ ID NO: 1:
			AC GACAATAAAA ATCAACTCAT TGCTGACTTG GGTTCTGAAC GCCGCGTCAA 60
			CT AATGATATTC CCACAGATTT GGTTAAGGCA ATCGTTTCTA TCGAAGACCA 120
			TC GACCACAGGG GGATTGATAC CATCCGTATC CTGGGAGCTT TCTTGCGCAA 180
	75	TCTGCAAA	SC AATTCCCTCC AAGGTGGATC AACTCTCACC CAACAGTTGA TTAAGTTGAC 240

PATENT APPLICATION: US/09/765,272A

DATE: 02/20/2003 TIME: 16:02:52

Input Set: A:\340P2C2 Substitute Sequence List 14 Feb 2003.txt
Output Set: N:\CRF4\02202003\I765272A.raw

77 TTACTTTTCA ACTTCGACTT CCGACCAGAC TATTTCTCGT AAGGCTCAGG AAGCTTGGTT	300										
79 AGCGATTCAG TTAGAACAAA AAGCAACCAA GCAAGAAATC TTGACCTACT ATATAAATAA	360										
81 GGTCTACATG TCTAATGGGA ACTATGGAAT GCAGACAGCA GCTCAAAACT ACTATGGTAA	420										
83 AGACCTCAAT AATTTAAGTT TACCTCAGTT AGCCTTGCTG GCTGGAATGC CTCAGGCACC	480										
85 AAACCAATAT GACCCCTATT CACATCCAGA AGCAGCCCAA GACCGCCGAA ACTTGGTCTT	540										
87 ATCTGAAATG AAAAATCAAG GCTACATCTC TGCTGAACAG TATGAGAAAG CAGTCAATAC	600										
89 ACCAATTACT GATGGACTAC AAAGTCTCAA ATCAGCAAGT AATTACCCTG CTTACATGGA	660										
91 TAATTACCTC AAGGAAGTCA TCAATCAAGT TGAAGAAGAA ACAGGCTATA ACCTACTCAC	720										
93 AACTGGGATG GATGTCTACA CAAATGTAGA CCAAGAAGCT CAAAAACATC TGTGGGATAT	780										
95 TTACAATACA GACGAATACG TTGCCTATCC AGACGATGAA TTGCAAGTCG CTTCTACCAT	840										
97 TGTTGATGTT TCTAACGGTA AAGTCATTGC CCAGCTAGGA GCACGCCATC AGTCAAGTAA	900										
99 TGTTTCCTTC GGAATTAACC AAGCAGTAGA AACAAACCGC GACTGGGGAT CAACTATGAA	960										
101 ACCGATCACA GACTATGCTC CTGCCTTGGA GTACGGTGTC TACGATTCAA CTGCTACTAT	1020										
103 CGTTCACGAT GAGCCCTATA ACTACCCTGG GACAAATACT CCTGTTTATA ACTGGGATAG	1080										
105 GGGCTACTTT GGCAACATCA CCTTGCAATA CGCCCTGCAA CAATCGCGAA ACGTCCCAGC	1140										
107 CGTGGAAACT CTAAACAAGG TCGGACTCAA CCGCGCCAAG ACTTTCCTAA ATGGTCTAGG	1200										
109 AATCGACTAC CCAAGTATTC ACTACTCAAA TGCCATTTCA AGTAACACAA CCGAATCAGA	1260										
111 CAAAAAATAT GGAGCAAGTA GTGAAAAGAT GGCTGCTGCT TACGCTGCCT TTGCAAATGG	1320										
113 TGGAACTTAC TATAAACCAA TGTATATCCA TAAAGTCGTC TTTAGTGATG GGAGTGAAAA	1380										
115 AGAGTTCTCT AATGTCGGAA CTCGTGCCAT GAAGGAAACG ACAGCCTATA TGATGACCGA	1440										
117 CATGATGAAA ACAGTCTTGA CTTATGGAAC TGGACGAAAT GCCTATCTTG CTTGGCTCCC	1500										
119 TCAGGCTGGT AAAACAGGAA CCTCTAACTA TACAGACGAG GAAATTGAAA ACCACATCAA	1560										
121 GACCTCTCAA TTTGTAGCAC CTGATGAACT ATTTGCTGGC TATACGCGTA AATATTCAAT	1620										
123 GGCTGTATGG ACAGGCTATT CTAACCGTCT GACACCACTT GTAGGCAATG GCCTTACGGT	1680										
125 CGCTGCCAAA GTTTACCGCT CTATGATGAC CTACCTGTCT GAAGGAAGCA ATCCAGAAGA	1740										
127 TTGGAATATA CCAGAGGGC TCTACAGAAA TGGAGAATTC GTATTTAAAA ATGGTGCTCG	1800										
131 ATCAGATAGT TCAACTTCAC AGTCTAGCTC AACCACTCCA AGCACAAATA ATAGTACGAC	1860 1920										
133 TACCAATCCT AACAATAATA CGCAACAATC AAATACAACC CCTGATCAAC AAAATCAGAA	1980										
135 TCCTCAACCA GCACAACCA	1999										
137 (2) INFORMATION FOR SEQ ID NO: 2:											
139 (i) SEQUENCE CHARACTERISTICS:											
140 (A) LENGTH: 666 amino acids											
141 (B) TYPE: amino acid											
142 (C) STRANDEDNESS: single											
143 (D) TOPOLOGY: linear											
145 (ii) MOLECULE TYPE: protein											
148 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:											
150 Lys Ile Tyr Asp Asn Lys Asn Gln Leu Ile Ala Asp Leu Gly Ser Glu											
151 1 5 10 15											
153 Arg Arg Val Asn Ala Gln Ala Asn Asp Ile Pro Thr Asp Leu Val Lys											
153 Arg Arg Var Ash Ara of Ara Ash A											
156 Ala Ile Val Ser Ile Glu Asp His Arg Phe Phe Asp His Arg Gly Ile											
150 Ala lie val sel lie Glu Asp his Alg the lie Asp his hig Gly lie 157 35 40 45											
157 Asp Thr Ile Arg Ile Leu Gly Ala Phe Leu Arg Asn Leu Gln Ser Asn											
160 50 55 60											
162 Ser Leu Gln Gly Gly Ser Thr Leu Thr Gln Gln Leu Ile Lys Leu Thr											
163 65 70 75 80											
165 Tyr Phe Ser Thr Ser Asp Gln Thr Ile Ser Arg Lys Ala Gln											
100 TAL THE DEL THE DEL THE DEL Wash GILL THE DEL WIR DAY WIR GILL											

PATENT APPLICATION: US/09/765,272A

DATE: 02/20/2003 TIME: 16:02:52

Input Set : A:\340P2C2 Substitute Sequence List 14 Feb 2003.txt
Output Set: N:\CRF4\02202003\I765272A.raw

Output	JCC.	11. (CIA 4 (02202005 (110521211.14#

166					85					90					95	
168	Glu	Ala	Trp	Leu	Ala	Ile	Gln	Leu	Glu	Gln	Lys	Ala	Thr	Lys	Gln	Glu
169			•	100					105		-			110		
171	Ile	Leu	Thr	Tvr	Tvr	Ile	Asn	Lvs	Val	Tvr	Met	Ser	Asn	Gly	Asn	Tyr
172			115	- 1	1			120		2			125	-		_
174				Thr	Ala	Ala	Gln	Asn	Tvr	Tvr	Glv	Lvs	Asp	Leu	Asn	Asn
175	1	130					135		-2 ·	-1		140				
177	Len		Len	Pro	Gln	Leu		Leu	Leu	Ala	Glv		Pro	Gln	Ala	Pro
178	145	001	200		0	150					155					160
180		Gln	Tur	Asp	Pro	Tyr	Ser	His	Pro	Glu		Ala	Gln	Asp	Ara	
181	11011	02.11	- <u>y</u> -	1101	165	- 1 -	001	1110		170		1110	0	P	175	5
183	Aen	Ι.Δ11	Val	T.011		Glu	Met	Lvs	Asn		Glv	Tur	Tle	Ser		Glu
184	nsii	пеп	Val	180	SCI	OIU	ricc	цуз	185	OIII	Ory	1 9 1	110	190	1114	Olu
	Cln	т	Clu		717	Val	Λen	Thr		Tla	Thr	Aen	Glv		Gln	Ser
186	GIII	тут	195	гуз	мта	vaı	ASII	200	110	116	1111	изр	205	пси	OLII	JCI
187	T 0.11	T		7) 1 -	C02	7 an	Ф		717	ጥ፣፣ም	Mot	7 cn		Туг	LOU	Luc
189	ьeu	_	ser	нта	ser	Asn	215	PIO	нта	тут	met	220	ASII	тут	пец	цуз
190	C1	210	т1.	7	C1-	W-1		C1	C1	Пhъ	C1		Λαη	Tou	T OU	Пhr
192		vai	тте	ASII	GTII	Val	GIU	GIU	GIU	TIII	_	ıyı	ASII	пеп	пец	
193	225	~1	N - 1	70	77 - 7	230	ml	7	17-3	7	235	C1	71.	C1 ~	T	240
195	Thr	GTÀ	мет	Asp		Tyr	THE	ASII	۷ат		GIII	GIU	Ald	GIII		птѕ
196	-		70	- 1	245	70	m1	71	C1	250	17- 1	71 -	mi-	Dece	255	7 ~~
198	Leu	Trp	Asp		Tyr	Asn	Tnr	Asp		Tyr	vaı	Ата	Tyr		Asp	ASP
199	~ 1	_	~ 3	260	2.1	•	m1	~ 1 .	265	70	777	0	70	270	T	77-3
201	GLu	Leu		Val	Ala	Ser	Thr		Val	Asp	val	Ser		GLY	гаг	vaı
202			275	_	- 1		_	280	~ 3	_	•		285	a	DI:	01
204	He		GIn	Leu	GLy	Ala		His	GIn	Ser	Ser		vaı	Ser	Pne	GTÀ
205		290					295	_	_	_	_	300	_			-
207		Asn	GIn	Ala	Val	Glu	Thr	Asn	Arg	Asp		GTA	Ser	Thr	мет	
208	305			_	_	310	_		-	~ 1	315	0 1	1	m.	70	320
210	Pro	He	'l'hr	Asp		Ala	Pro	Ala	Leu		Tyr	GTY	vaı	Tyr		Ser
211					325		_	~ 1	_	330	_	_	-	6 3	335	_
213	Thr	Ala	Thr		Val	His	Asp	Glu		Tyr	Asn	Tyr	Pro		Thr	Asn
214		_	•	340	_	_	_	·	345	_	D 1	6 3	_	350	m1	`_
216	Thr	Pro		Tyr	Asn	Trp	Asp	-	GLy	Tyr	Phe	Gly		шe	Thr	Lеu
217			355				_	360	_		_		365	~ 1		_
219	Gln	_	Ala	Leu	Gln	Gln		Arg	Asn	Val	Pro		Val	Glu	Thr	Leu
220		370					375					380				
222		Lys	Val	Gly	Leu	Asn	Arg	Ala	Lys	Thr		Leu	Asn	Gly	Leu	
223	385					390					395					400
225	Ile	Asp	Tyr	Pro		Ile	His	Tyr	Ser		Ala	Ile	Ser	Ser		Thr
226					405					410					415	
228	Thr	Glu	Ser	Asp	Lys	Lys	Tyr	Gly	Ala	Ser	Ser	Glu	Lys	Met	Ala	Ala
229				420					425					430		
231	Ala	Tyr	Ala	Ala	Phe	Ala	Asn	Gly	Gly	Thr	Tyr	Tyr	Lys	Pro	Met	Tyr
232			435					440					445			
234	Ile	His	Lys	Val	Val	Phe	Ser	Asp	Gly	Ser	Glu	Lys	Glu	Phe	Ser	Asn
235		450					455					460				
237	Val	Gly	Thr	Arg	Ala	Met	Lys	Glu	Thr	Thr	Ala	Tyr	Met	Met	Thr	Asp
238	465	-		-		470	•				475					480

PATENT APPLICATION: US/09/765,272A

DATE: 02/20/2003 TIME: 16:02:52

Input Set: A:\340P2C2 Substitute Sequence List 14 Feb 2003.txt Output Set: N:\CRF4\02202003\I765272A.raw

								_			~-	_	_	~ 1	_	-	
240	Met	Met	Lys	Thr		Leu	Thr	Tyr	GLy		Gly	Arg	Asn	Ala		Leu	
241					485					490					495		
243	Ala	Trp	Leu		Gln	Ala	Gly	Lys		Gly	Thr	Ser	Asn		Thr	Asp	
244				500					505					510			
246	Glu	Glu	Ile	Glu	Asn	His	Ile	Lys	Thr	Ser	Gln	Phe		Ala	Pro	Asp	
247			515					520					525				
249	Glu	Leu	Phe	Ala	Gly	Tyr	Thr	Arg	Lys	Tyr	Ser	Met	Ala	Val	Trp	Thr	
250		530					535					540					
252	Gly	Tyr	Ser	Asn	Arg	Leu	Thr	Pro	Leu	Val	Gly	Asn	Gly	Leu	Thr	Val	
253	545					550					555					560	
255	Ala	Ala	Lys	Val	Tyr	Arg	Ser	Met	Met	Thr	Tyr	Leu	Ser	Glu	Gly	Ser	
256					565					570					575		
258	Asn	Pro	Glu	Asp	Trp	Asn	Ile	Pro	Glu	Gly	Leu	Tyr	Arg	Asn	Gly	Glu	
259				580					585					590			
261	Phe	Val	Phe	Lys	Asn	Gly	Ala	Arg	Ser	Thr	Trp	Asn	Ser	Pro	Ala	Pro	
262			595					600					605				
264	Gln	Gln	Pro	Pro	Ser	Thr	Glu	Ser	Ser	Ser	Ser	Ser	Ser	Asp	Ser	Ser	
265		610					615					620					
267	Thr	Ser	Gln	Ser	Ser	Ser	Thr	Thr	Pro	Ser	Thr	Asn	Asn	Ser	Thr	Thr	
268	625					630					635		•			640	
270	Thr	Asn	Pro	Asn	Asn	Asn	Thr	Gln	Gln	Sér	Asn	Thr	Thr	Pro	Asp	Gln	
271					645					650					655		
273	Gln	Asn	Gln	Asn	Pro	Gln	Pro	Ala	Gln	Pro							
273 Gln Asn Gln Asn Pro Gln Pro Ala Gln Pro 274 660 665																	
276 (2) INFORMATION FOR SEQ ID NO: 3:																	
278																	
279	279 (A) LENGTH: 1714 base pairs																
280	280 (B) TYPE: nucleic acid																
281																	
282		(D) TO	POLO	GY: :	linea	ar										
286	(xi)	SEQ	UENCI	E DES	SCRI	OITS	N: S1	EQ II	ON C	: 3:							
288	AAATTACA	AT A	CGGA	CTAT	G AA	rtga(CCTC	TGG	AGAA?	AAA :	TTAC	CTCT	rc c)AAA1	GAGA!	Γ	60
290	TTCAGGTT	AC A	CTTA:	TATT	G GA	rata:	ГСАА	AGA(GGGA	AAA A	ACGA	CTTC:	rg A	GTCT	GAAG'	$f \Gamma$	120
292	AAGTAATC	AA AA	AGAG:	TTCA	G TTO	GCCA(CTCC	TACA)AAA/	CAA (CAAA	AGGT	GG A	TATA	AATG	Γ	180
	TACACCGA																240
	TTCTTCAA																300
	CAATCCAA																360
300	GAATCTAGA	AA AA	CGAA	GAAA	G AG	GAGA	AGAT	TTC	rcca/	AAA (GAAA	AGAC:	rg go	GGTAA	ATA	C	420
302	ATTAAATC	CA C	AGGA:	TGAA(G TT	TAT	CAGG	TCA	ATTGA	AAC A	AAAC	CTGA	AC TO	CTTA:	ratco	G	480
304	TGAGGAAA	CT A	TGGA(GACA	AAA A	ATAGA	TTTP	TCA	AGAA	GAA A	ATTC	AAGA?	AA A	rcct(SATT	Г	540
306	AGCTGAAG	GA AG	CTGTA	AAGA	G TAA	AAAC	AAGA	AGG.	CAAA1	TTA (GGTA	AGAA?	AG T	rgaa <i>i</i>	ATCG:	Г	600
	CAGAATAT																660
	TGCGCCTA																720
	ACCTGAGAG																780
	TCAGCCTG																840
	ATTACCCG																900
	TACTGTGG																960
	TAATATTG																1020
322	AAAAACTG	AA G	AAGT	TCCA(G TAA	AAAC	CAAC	AGA	\GAA/	ACA (CCAG	raaar	rc ca	TAAP	GAAG(G	1080

PATENT APPLICATION: US/09/765,272A

DATE: 02/20/2003 TIME: 16:02:52

Input Set : A:\340P2C2 Substitute Sequence List 14 Feb 2003.txt
Output Set: N:\CRF4\02202003\I765272A.raw

140 140																		
1260 1270	324	TACTACAG	AA G	GAAC	CTCA	A TTO	CAAG	AAGC	AGA	TAAA	CCA (GTTC	AACC!	rg ca	AGAA(GAAT	C	1140
1320 CTCTARARAT GARATTCAG ARARACTGT AGAGARGTT CCAGTRARTC CARATGRAGG 1320 332 CACAGTAGAA GGTACCTCAR ATCAGARAC AGRARACAAC AGAGACAAC CAGAGAGAAC 1340 340	326	AACAACGA	T TA	CAGA	GAAA	G TA	rcac(CAGA	TAC	ATCTA	AGC 2	AAAA	ATAC	rg go	GGAA(GTGT(C	1200
332 CACACTAGAA GETACCTCAA ATCAAGAAAC AGAAAAACCA GTTCAACCTG CAGAAGAAAC 334 ACAAACAAAC TCTGGGAAAA TAGCTAACGA AAATACTGGA GAAGTATCCA ATAAACCTA 346 CAGTTCAAAAC CCACCAGTTG AAGAATCAAAA TCAACCAGAAC AAAAAACCGAAC CAGACACAAA 1500 338 ACCAGAAAAT TCAGGTAATA CAACAATCAAA TCAACCAGAACAAA ACAACCACCACAAA 1500 338 ACCAGAAAAT TCAGGTAATA CAACATCAGA GAATGGACAA AAAAACCGAA CACCACACAACA 1500 340 CAGGAAATTCA ACTGAGGATG TTCAACCGA ATCAAACCAC TCCAATTCAA ATGAAACGA 1620 341 ACTTGAATTAA CAAGAAAATG AACTAGACCC TGATAAAAAG GTAGAAGAAC CAGAGAAAAC 1680 344 ACTTGAATTAA ACAAGAACTGTT CCCACCTAGA GTTA 1714 348 (i) SEQUENCE CHARACTERISTICS: 349 (A) LENGTH: 571 amino acid 351 (C) STRANDENNESS: single (B) TYPE: amino acid (B) TYPE: amino acid (B) TYPE: protein 357 (Xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4: 360 1	328	CAGTAATC	CT A	GTGA'	TTCG	A CA	ACCT	CAGT	TGG	AGAA:	rca i	AATAA	AACC	AG AZ	ACATA	AATG	A	1260
334 ACANACAAAC TCTGGGGAAAA TAGGTAACGA AAATACTGGA GAAGTATCCA ATACACCAGA	330	CTCTAAAA	AT G	'AAAA	TTCA	G AA	AAAA	CTGT	AGA	AGAA(GTT (CCAG'	raaa'	rc ca	TAA	GAAG	G	1320
334 ACANACAAAC TCTGGGGAAAA TAGGTAACGA AAATACTGGA GAAGTATCCA ATACACCAGA	332															1380		
338 ACCAGAAATT TCAGGTAATA CAACATCACA GAATGGACAA ACGGAACCA TCAAA ATGGAACCA CGGAAATTCA ACTGAGGATG TTTCAACCGA ATCAAACACA TCCAAATTCAA ATGGAAACGA 1620 42 ACATGAAATTAA CAACAATGTTC CCACCTACA GTTA 1680 344 ACTTGAATTA ACACAATGTTC CCACCTACA GTTA 1714 46 (2) INFORMATION FOR SEQ ID NO: 4:	334	ACAAACAAAC TCTGGGAAAA TAGCTAACGA AAATACTGGA GAAGTATCCA ATAAACCTAG														1440		
338 ACCAGARAAT TCAGGTAATA CAACATCACA GAATGGACAA ACCATCAAA ATGGAAACCA 340 CGGAAATTCAA ACTGAGGATG TTTCAACCGA ATCAAACACA TCCAAATCAAA ATGGAAACCA 342 AGAAATTAAA CAAGAAAATG AACTAGCCC TGATAAAAAAG GTAGAAGAAC CAGAGAAAAC 343 ACTGGAATTCA ACGAAAATG AACTAGCCC TGATAAAAAAG GTAGAAGAAC CAGAGAAAAC 348 (1) SEQUENCE CHARACTERISTICS: 349	336	6 TGATTCAAAA CCACCAGTTG AAGAATCAAA TCAACCAGAA AAAAACGGAA CTGCAACAAA															1500	
340 GGARATTCA ACTGAGGATG TTTCAACCCA ATCAACCA TCCAATTCAA ATGGARACCA 1680 342 AGAAATTAAA CAAGAAAATG AACTAGACCC TGATAAAAAG GTAGAAGAAC CAGAGAAAAC 1680 1714 1680 1714 1680 1714 1680 1714 1680 1714 17																		1560
344 ACTIGARITA AGAGARARIG ARCTACACCC TGATARARAG GTAGARGAC CAGAGARACC 1714																		1620
ACTTGAATTA AGAAATGTTT CCGACCTAGA GTTA 1714 346 (2) INFORMATION FOR SEQ ID NO: 4:																		1680
(2) INFORMATION FOR SEQ ID NO: 4: (348 (1) SEQUENCE CHARACTERISTICS: (A) LENGTH: 571 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: protein (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4: (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4: 357 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4: 368 (1) Tyr Asn Thr Asp Tyr Glu Leu Thr Ser Gly Glu Lys Leu Pro Leu 360 (1) 5 (2) 20 25 30 365 (2) Pro Lys Glu Ile Ser Gly Tyr Thr Tyr Ile Gly Tyr Ile Lys Glu Gly 366 (3) 20 25 30 367 (4) Ser Thr Ser Glu Ser Glu Val Ser Asn Gln Lys Ser Ser Val Ala 368 (3) 40 45 369 50 50 70 75 80 371 Val Asp His Pro Ser Thr Val Gln Ala Ile Gln Glu Gln Thr Pro Val 372 (65 70 75 80 373 (8) Ser Ser Thr Lys Pro Thr Glu Val Gln Val Val Glu Lys Pro Phe Ser 375 85 90 95 376 Thr Glu Leu Ile Asn Pro Arg Lys Glu Glu Lys Gln Ser Ser Asp Ser 377 Thr Glu Leu Ile Asn Pro Arg Lys Glu Glu Lys Gln Ser Ser Asp Ser 378 100 Gln Glu Gln Leu Ala Glu His Lys Asn Leu Glu Thr Lys Lys Glu Glu 380 Gln Glu Gln Leu Ala Glu His Lys Asn Leu Glu Thr Lys Lys Glu Glu 381 130 135 140 382 Asp Glu Val Leu Ser Gly Glu Lys Thr Gly Val Asn Thr Leu Asn Pro Gln 384 130 135 150 125 385 160 170 175 396 Glu Glu Thr Met Glu Thr Lys Ile Asp Phe Gln Glu Glu Lu Leu Tyr Arg 387 145 150 170 175 398 Glu Glu Thr Met Glu Thr Lys Ile Asp Phe Gln Glu Glu Lu Lys Ile Glu Glu 399 180 180 195 180 185 180 190 398 Glu Val Ser Arg Glu Ile Val Ser Thr Ser Thr Thr Thr Leu Asn Lys Glu 399 190 210 225 398 Glu Val Ser Arg Glu Ile Val Ser Thr Ser Thr Thr Thr Leu Asn Lys Glu 399 210 215 220 401 Arg Ile Val Glu Lys Gly Thr Lys Lys Thr Gln Val Ile Lys Glu Glu																		1714
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 571 amino acids (B) TYPE: amino acids (C) STRANDEDNESS: single (C) STRANDEDNESS: single (D) TOPOLOGY: linear (Xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4: 359 Asn Tyr Asn Thr Asp Tyr Glu Leu Thr Ser Gly Glu Lys Leu Pro Leu 15 10 15 360 1 5 10 15 362 Pro Lys Glu Ile Ser Gly Tyr Thr Tyr Ile Gly Tyr Ile Lys Glu Gly 363 20 25 365 Lys Thr Thr Ser Glu Ser Glu Val Ser Asn Gln Lys Ser Ser Val Ala 366 35 40 40 368 Thr Pro Thr Lys Gln Gln Lys Val Asp Tyr Asn Val Thr Pro Asn Phe 369 50 55 60 371 Val Asp His Pro Ser Thr Val Gln Ala Ile Gln Glu Gln Thr Pro Val 372 65 70 374 Ser Ser Thr Lys Pro Thr Glu Val Gln Val Val Glu Lys Pro Phe Ser 375 85 90 377 Thr Glu Leu Ile Asn Pro Arg Lys Glu Glu Lys Gln Ser Ser Asp Ser 378 100 105 380 Gln Glu Gln Leu Ala Glu His Lys Asn Leu Glu Thr Lys Lys Glu Glu 381 115 120 382 Lys Ile Ser Pro Lys Glu Lys Thr Gly Val Asn Thr Leu Asn Pro Gln 384 130 135 140 385 145 150 165 387 165 160 388 Glu Glu Thr Met Glu Thr Lys Ile Asp Phe Glu Glu Glu Ile Gln Glu 389 Glu Glu Thr Met Glu Thr Lys Ile Asp Phe Glu Glu Glu Ile Gln Glu 390 165 170 390 165 170 391 180 180 180 180 180 190 392 Asn Pro Asp Leu Ala Glu Gly Thr Val Arg Val Lys Glu Glu Ile Gln Glu 395 Leu Gly Lys Lys Val Glu Ile Val Arg Val Lys Glu Glu Glu 396 Glu Val Ser Arg Glu Ile Val Ser Thr Ser Thr Thr Ala Pro Ser Pro 399 210 225 401 Arg Ile Val Glu Lys Gly Thr Lys Lys Thr Gln Val Ile Lys Glu Gln 399 Glu Val Ser Arg Glu Ile Val Ser Thr Ser Thr Thr Ala Pro Ser Pro 399 210 225		346 (2) INFORMATION FOR SEQ ID NO: 4:																
(A) LENGTH: 571 amino acids (B) TYPE: amino acid (C) STRANDENNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: protein (iii) MOLECULE TYPE:																		
350																		
C STRANDEDNESS: single (D) TOPOLOGY: linear (D) TOPOLOGY: linear (E) TYPE: protein (E) SEQUENCE DESCRIPTION: SEQ ID NO: 4: (E) SEQ ID S																		
352		i de la companya de																
354		· · · · · · · · · · · · · · · · · · ·																
357		(ii)																
Ash Tyr Ash Thr Asp Tyr Glu Leu Thr Ser Gly Glu Lys Leu Pro Leu 15									EO II	ON C	: 4:							
1												Glv	Glu	Lys	Leu	Pro	Leu	
Pro Lys Glu Ile Ser Gly Tyr Thr Tyr Ile Gly Tyr Ile Lys Glu Gly 363 20		_	- 1 -			_	- 1 -					-4		-				
363 20 25 30 365 Lys Thr Thr Ser Glu Ser Glu Val Ser Asn Gln Lys Ser Ser Val Ala 36 366 35 40 45 368 Thr Pro Thr Lys Gln Gln Lys Val Asp Tyr Asn Val Thr Pro Asn Phe 36 369 50 55 60 371 Val Asp His Pro Ser Thr Val Gln Ala Ile Gln Glu Gln Thr Pro Val 372 372 65 70 70 75 80 374 Ser Ser Thr Lys Pro Thr Glu Val Gln Val Glu Lys Pro Pro Pro Ser 85 90 95 377 Thr Glu Leu Ile Asn Pro Arg Lys Glu Glu Lys Gln Ser Ser Asp Ser 100 105 110 380 Gln Glu Gln Leu Ala Glu His Lys Asn Leu Glu Thr Lys Lys Glu Glu 120 125 383 Lys Ile Ser Pro Lys Glu Lys Thr Gly Val Asn Thr Leu Asn Pro Gln 384 130 125 384 130 15 120 125 386 Asp Glu Val Leu Ser Gly Gln Leu Asn Lys Pro Glu Leu Leu Tyr Arg 140 387 145 150 155 160 389 Glu Glu Thr Met Glu Thr Lys Ile Asp Phe Gln Glu Glu Glu Glu Glu Glu 125 175 392			Lvs	Glu	Ile		Glv	Tvr	Thr	Tvr		Glv	Tvr	Ile	Lys	Glu	Gly	
Second			-1-				1	4				-	-				-	
366 368 Thr Pro Thr Lys Gln Gln Lys Val Asp Tyr Asn Val Thr Pro Asn Phe 50		Lvs	Thr	Thr		Glu	Ser	Glu	Val		Asn	Gln	Lvs	Ser	Ser	Val	Ala	
368 Thr Pro Thr Lys Gln Gln Lys Val Asp Tyr Asn Val Thr Pro Asn Phe 369 50 50 55 60 70 75 60 371 Val Asp His Pro Ser Thr Val Gln Ala Ile Gln Glu Glu Gln Thr Pro Val 70 75 80 372 65 70 75 80 374 Ser Ser Thr Lys Pro Phe Ser 85 90 95 377 Thr Glu Leu Ile Asn Pro Arg Lys Glu Glu Lys Gln Ser Ser Asp Ser 95 378 100 105 110 380 Gln Glu Gln Leu Ala Glu His Lys Asn Leu Glu Thr Lys Lys Glu Glu 125 381 115 120 120 384 130 135 140 386 Asp Glu Val Leu Ser Pro Lys Glu Lys Thr Gly Val Asn Thr Leu Asn Pro Gln 140 386 Asp Glu Val Leu Ser Gly Gln Leu Asn Lys Pro Glu Leu Leu Tyr Arg 150 387 145 150 150 150 150 389 Glu Glu Thr Met Glu Thr Lys Ile Asp Phe Gln Glu Glu Glu Leu Leu Tyr Arg 175 160 390 165 170 175 392 Asn Pro Asp Leu Ala Glu		-1-				- " -							-					
369 50 55 60 371 Val Asp His Pro Ser Thr Val Gln Ala Ile Gln Glu Gln Thr Pro Val 75 80 372 65 70 75 80 374 Ser Ser Thr Lys Pro Thr Glu Val Gln Val Val Glu Lys Pro Phe Ser 85 90 95 375 85 90 95 95 377 Thr Glu Leu Ile Asn Pro Arg Lys Glu Glu Lys Glu Lys Gln Ser Ser Asp Ser 100 110 130 110 130 380 Gln Glu Gln Leu Ala Glu His Lys Asn Leu Glu Thr Lys Lys Glu Glu 125 120 125 125 130 140 125 140 125 140 140 140 125 150 140 125 150 140 125 150 140		Thr	Pro		Lvs	Gln	Gln	Lvs	Val	Asp	Tvr	Asn	Val	Thr	Pro	Asn	Phe	
371 Val Asp His Pro Ser Thr Val Gln Ala Ile Gln Glu Gln Thr Pro Val 372 65 70 75 80 374 Ser Ser Thr Lys Pro Thr Glu Val Gln Val Gln Val Val Glu Lys Pro Phe Ser 85 90 95 375 Thr Glu Leu Ile Asn Pro Arg Lys Glu Glu Lys Gln Ser Ser Asp Ser 100 105 110 380 Gln Glu Gln Leu Ala Glu His Lys Asn Leu Glu Thr Lys Lys Glu Glu Sal 110 120 125 383 Lys Ile Ser Pro Lys Glu Lys Thr Gly Val Asn Thr Leu Asn Pro Gln 135 140 384 130 135 140 386 Asp Glu Val Leu Ser Gly Gln Leu Asn Lys Pro Glu Leu Leu Tyr Arg 387 145 150 155 160 389 Glu Glu Thr Met Glu Thr Lys Ile Asp Phe Gln Glu Glu Ile Gln Glu 175 175 392 Asn Pro Asp Leu Ala Glu Gly Thr Val Arg Val Lys Gln Glu Gly Lys 185 190 393 180 185 190 395 Leu Gly Lys Lys Val Glu Ile Val Arg Ile Phe Ser Val Asn Lys Glu 396 200 205 398 Glu Val Ser Arg Glu Ile Val Ser Thr Ser Thr Thr Ala Pro Ser Pro 399 210 220					-1-					- 1	_							
372 65 70 75 80 374 Ser Ser Thr Lys Pro Rhr Glu Val Glu Val Glu Val Glu Val Glu Lys Pro Phe Ser 85 85 90 95 375 Thr Glu Leu Ile Asn Pro Rrg Lys Glu Glu Lys Glu Glu Lys Gln Ser Ser Asp Ser 100 105 110 380 Gln Glu Gln Leu Ala Glu His Lys Asn Leu Glu Thr Lys Lys Glu Glu 381 115 120 383 Lys Ile Ser Pro Lys Glu Lys Thr Gly Val Asn Thr Leu Asn Pro Gln 130 135 140 386 Asp Glu Val Leu Ser Gly Gln Leu Asn Lys Pro Glu Leu Leu Tyr Arg 155 160 389 Glu Glu Thr Met Glu Thr Lys Ile Asp Phe Gln Glu Glu Ile Gln Glu 390 165 170 175 392 Asn Pro Asp Leu Ala Glu Gly Thr Val Arg 185 190 195 190 395 Leu Gly Lys Lys Val Glu Ile Val Arg Ile Phe Ser Val Asn Lys Glu 396 190 205 398 Glu Val Ser Arg Glu Ile Val Ser Thr Ser Thr Thr Ala Pro Ser Pro 205 205 398 Glu Val Ser Arg Glu Ile Val Ser Thr Lys Lys Thr Gln Val Ile Lys Glu Gln 399 Glu Val Ser Arg Glu Ile Val Ser Thr Ser Thr Thr Ala Pro Ser Pro 210 399 Leu Gly Val Glu Lys Gly Thr Lys Lys Thr Gln Val Ile Lys Glu Gln		Val		His	Pro	Ser	Thr	Val	Gln	Ala	Ile	Gln	Glu	Gln	Thr	Pro	Val	
374 Ser Ser Thr Lys Pro Thr Glu Val Glu Val Glu Lys Pro Phe Ser 375 Thr Glu Leu Ile Asn Pro Arg Lys Glu Glu Lys Glu Ser Ser Asp Ser 377 Thr Glu Leu Ile Asn Pro Arg Lys Glu Glu Lys Glu Lys Glu Glu Lys Glu Glu Lys Glu Glu Lys Glu Glu Lys Lys Leu Glu Int Lys Lys Int Lys Int Lys Int Lys Int Lys Int I			1															
375 Thr Glu Leu Ile Asn Pro Arg Lys Glu Glu Lys Gln Ser Ser Asp Ser 100 110			Ser	Thr	Lys	Pro	Thr	Glu	Val	Gln	Val	Val	Glu	Lys	Pro	Phe	Ser	
377 Thr Glu Leu Ile Asn Pro Arg Lys Glu Glu Lys Gln Ser Ser Asp Ser 100 100 105 110 Ser Asp Ser 110 380 Gln Glu Gln Leu Gln Leu Ala Glu His Lys Asn Leu Glu Thr Lys Lys Glu Glu 115 120 125 125 126 125 126 127 128 128 128 128 129 128 129 </td <td></td> <td></td> <td></td> <td></td> <td>-</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>-</td> <td></td> <td></td> <td></td> <td></td>					-									-				
378 380 380 381 381 381 383 383 384 384 386 386 386 387 387 387 387 387 389 389 380 380 380 380 380 380 380 380 380 380		Thr	Glu	Leu	Ile	Asn	Pro	Arg	Lys	Glu	Glu	Lys	Gln	Ser	Ser	Asp	Ser	
380 Gln Glu Gln Leu Ala Glu His Lys Asn Leu Glu Thr Lys Lys Glu Glu 381 115 120 125 383 Lys Ile Ser Pro Lys Glu Lys Thr Gly Val Asn Thr Leu Asn Pro Gln 384 130 5 135 5 140 386 Asp Glu Val Leu Ser Gly Gln Leu Asn Lys Pro Glu Leu Leu Tyr Arg 387 145 150 150 155 160 389 Glu Glu Thr Met Glu Thr Lys Ile Asp Phe Gln Glu Glu Ile Gln Glu 390 165 170 175 392 Asn Pro Asp Leu Ala Glu Gly Thr Val Arg Val Lys Gln Glu Gly Lys 393 180 185 190 395 Leu Gly Lys Lys Val Glu Ile Val Arg Ile Phe Ser Val Asn Lys Glu 396 Glu Val Ser Arg Glu Ile Val Ser Thr Ser Thr Thr Ala Pro Ser Pro 399 210 215 220 401 Arg Ile Val Glu Lys Gly Thr Lys Lys Thr Gln Val Ile Lys Glu Gln								_	-			-				_		
381 Lys Ile Ser Pro Lys Glu Lys Thr Gly Val Asn Thr Leu Asn Pro Gln 384 130 130 135 140 1		Gln	Glu	Gln	Leu	Ala	Glu	His	Lys	Asn	Leu	Glu	Thr	Lys	Lys	Glu	Glu	
383 Lys Ile Ser Pro Lys Glu Lys Thr Gly Val Asn Thr Leu Asn Pro Glu Leu Asn Pro Glu Leu Leu Thr Arg Arg Arg Pro Glu Leu Leu Thr Arg Arg Pro Glu Leu Leu Thr Arg Inch																		
384 130 135 140 386 Asp Glu Val Leu Ser Gly Gln Leu Asn Lys Pro Glu Leu Leu Tyr Arg 387 145 150 155 160 389 Glu Glu Thr Met Glu Thr Lys Ile Asp Phe Gln Glu Glu Ile Gln Glu 175 175 390 165 170 175 392 Asn Pro Asp Leu Ala Glu Gly Thr Val Arg Val Lys Gln Glu Gly Lys 190 393 180 185 190 395 Leu Gly Lys Lys Lys Val Glu Ile Val Arg Ile Phe Ser Val Asn Lys Glu 190 396 195 200 205 398 Glu Val Ser Arg Glu Ile Val Ser Thr Ser Thr Thr Ala Pro Ser Pro 399 210 215 200 220 401 Arg Ile Val Glu Lys Gly Thr Lys Lys Thr Gln Val Ile Lys Glu Gln	383	Lys	Ile	Ser	Pro	Lys	Glu	Lys	Thr	Gly	Val	Asn	Thr	Leu	Asn	Pro	Gln	
387 145 150 155 160 389 Glu Glu Thr Met Glu Thr Lys Ile Asp Phe Gln Glu Glu Ile Gln Glu 165 170 175 390 Asn Pro Asp Leu Ala Glu Gly Thr Val Arg Val Lys Gln Glu Gly Lys 175 392 Asn Pro Asp Leu Ala Glu Gly Thr Val Arg Val Lys Gln Glu Gly Lys 393 180 185 190 395 Leu Gly Lys Lys Lys Val Glu Ile Val Arg Ile Phe Ser Val Asn Lys Glu 190 396 195 200 205 398 Glu Val Ser Arg Glu Ile Val Ser Thr Ser Thr Thr Ala Pro Ser Pro 399 210 215 220 401 Arg Ile Val Glu Lys Gly Thr Lys Lys Thr Gln Val Ile Lys Glu Gln	384	-																
389 Glu Glu Thr Met Glu Thr Lys Ile Asp Phe Gln Glu Glu Ile Gln Glu 390	386	Asp	Glu	Val	Leu	Ser	Gly	Gln	Leu	Asn	Lys	Pro	Glu	Leu	Leu	Tyr	Arg	
390 165 170 175 392 Asn Pro Asp Leu Ala Glu Gly Thr Val Arg Val Lys Gln Glu Gly Lys 393 180 185 190 395 Leu Gly Lys Lys Val Glu Ile Val Arg Ile Phe Ser Val Asn Lys Glu 396 195 200 205 398 Glu Val Ser Arg Glu Ile Val Ser Thr Ser Thr Thr Ala Pro Ser Pro 399 210 215 220 401 Arg Ile Val Glu Lys Gly Thr Lys Lys Thr Gln Val Ile Lys Glu Gln	387	145					150					155					160	
392 Asn Pro Asp Leu Ala Glu Gly Thr Val Arg Val Lys Gln Glu Gly Lys 393 180 185 190 395 Leu Gly Lys Lys Val Glu Ile Val Arg Ile Phe Ser Val Asn Lys Glu 396 200 205 398 Glu Val Ser Arg Glu Ile Val Ser Thr Ser Thr Thr Ala Pro Ser Pro 399 210 215 220 401 Arg Ile Val Glu Lys Gly Thr Lys Lys Thr Gln Val Ile Lys Glu Gln	389	Glu	Glu	Thr	Met	Glu	Thr	Lys	Ile	Asp	Phe	Gln	Glu	Glu	Ile	Gln	Glu	
393 180 185 190 395 Leu Gly Lys Lys Val Glu Ile Val Arg Ile Phe Ser Val Asn Lys Glu 396 195 200 205 398 Glu Val Ser Arg Glu Ile Val Ser Thr Ser Thr Thr Ala Pro Ser Pro 399 210 215 220 401 Arg Ile Val Glu Lys Gly Thr Lys Lys Thr Gln Val Ile Lys Glu Gln	390					165		_			170					175		
393 180 185 190 395 Leu Gly Lys Lys Val Glu Ile Val Arg Ile Phe Ser Val Asn Lys Glu 396 195 200 205 398 Glu Val Ser Arg Glu Ile Val Ser Thr Ser Thr Thr Ala Pro Ser Pro 399 210 215 220 401 Arg Ile Val Glu Lys Gly Thr Lys Lys Thr Gln Val Ile Lys Glu Gln	392	Asn	Pro	Asp	Leu	Ala	Glu	Gly	Thr	Val	Arg	Val	Lys	Gln	Glu	Gly	Lys	
396 195 200 205 398 Glu Val Ser Arg Glu Ile Val Ser Thr Ser Thr Thr Ala Pro Ser Pro 399 210 215 220 401 Arg Ile Val Glu Lys Gly Thr Lys Lys Thr Gln Val Ile Lys Glu Gln				_														
396 195 200 205 398 Glu Val Ser Arg Glu Ile Val Ser Thr Ser Thr Thr Ala Pro Ser Pro 399 210 215 220 401 Arg Ile Val Glu Lys Gly Thr Lys Lys Thr Gln Val Ile Lys Glu Gln		Leu	Gly	Lys		Val	Glu	Ile	Val	Arg	Ile	Phe	Ser	Val	Asn	Lys	Glu	
Glu Val Ser Arg Glu Ile Val Ser Thr Ser Thr Thr Ala Pro Ser Pro 210 215 220 Arg Ile Val Glu Lys Gly Thr Lys Lys Thr Gln Val Ile Lys Glu Gln			-	_	-					-						•		
399 210 215 220 401 Arg Ile Val Glu Lys Gly Thr Lys Lys Thr Gln Val Ile Lys Glu Gln		Glu	Val	Ser	Arg	Glu	Ile	Val	Ser	Thr	Ser	Thr	Thr	Ala	Pro	Ser	Pro	
401 Arg Ile Val Glu Lys Gly Thr Lys Lys Thr Gln Val Ile Lys Glu Gln					_													
		Arq	Ile	Val	Glu	Lys	Gly	Thr	Lys	Lys	Thr	Gln	Val	Ile	Lys	Glu	Gln	
	402	225				-	_		-	_					•			